Interval-valued Feature Selection

Albert Samà, Francisco J. Ruiz, Andreu Català & Cecilio Angulo

CETpD -Technical Research Centre for Dependency Care and Autonomous Living
Neàpolis Building. Rambla de l'Exposició, 59-69. 08800 Vilanova i la Geltrú
email:albert.sama@gmail.com

Abstract

In this paper we introduce the use of interval variables in classification problems of time series signals. By introducing the concept of interval kernel as similarity measure among intervals, modifications for some well-known feature selection methods are developed in order to apply these methods to select the most relevant interval variables. A comparison against standard point attributes feature selection (Relief and FSDD) is made for purposes of validation .

1 Introduction

Interval-valued data is usually thought as a form to deal with uncertain data, but it is also a good way to aggregate a large amount of individual data into a smaller quantity of information which is more easily understandable and manageable. In this paper we propose the use of intervals to describe a signal for recognition tasks.

Signal recognition is a useful area in domains such as biomedical signals, continuous systems diagnosis and data mining in temporal databases. In signal recognition each pattern consists of one or more time series segments (figure 1a), that is, a set of values measured over time. From the original time series, others time series can be obtained by differentiating this series. Also, whether a single pattern is formed by a set of signals captured by several sensors, times series consisting of differences and ratios among these signals can also be considered.

The more simple representation of a pattern is considering the mean values of each of these series but, it is well known that measures of central tendency are not enough to accurately describe a data set. It is at least necessary to describe the variability or the data dispersion. Therefore, interval features is a good way to characterize signals (figure 1b).

In the case that each pattern consists of several signals, a huge amount of interval features can be considered (the original signals, increments, differences and ratios). For this reason, some feature selection process should be performed in order to choose a small subset of features that ideally is necessary and sufficient to describe the target concept. Too few features lead to unsatisfactory results due to lack of information and too many features increase the computational cost of the processing task and decrease the understanding of the recognition process.

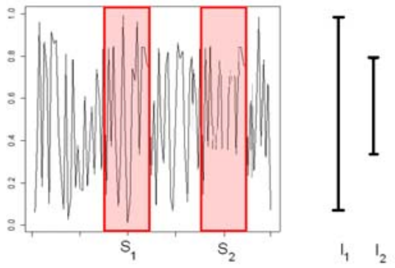

Figure 1. Two single patterns S_1 and S_2 (a) and its representation by means of intervals I_1 and I_2 (b).

Some feature selection methods are based on attribute weighting, which consists on assigning a relevance to each attribute and then selecting those with higher values. This strategy allows to avoid the cost of a combinatorial search between feature subsets. Most of weighting algorithms use measures of similarity or dissimilarity [4] [6] since maximization between instances of different class is aimed. Hence, if interval variables are used, specific similarity and dissimilarity measures should be considered in order to obtain efficient feature selection process.

In this paper we propose the use of interval features to describe a signal in order to solve classification problems and also compare the use of these interval features in two feature selection methods: Relief and Feature selection based on distance discriminant (FSDD) against standard point attributes. To this end we introduce the concept of interval kernel that can be considered as a measure of similarity.

The paper is organized as follows: In Section 2 some representative feature selection methods are introduced. Section 3 is devoted to the concept of interval kernel. In this section it is justified that this kernel can be considered as a similarity measure suitable to use in standard feature selection methods. Section 4 is devoted to the necessary modification of feature selection methods introduced in section 2 for using these methods with interval-valued data. Section 5 provides an example using the methodology introduced in a classification task with a dataset from UCI Machine Learning. The last section collects some conclusions and also some additional suggestions for future work are given.

2 Feature selection

Dimensionality reduction methods can be broadly classified into two groups: feature extraction (such as principal component analysis (PCA) or linear discriminant analysis (LDA) and feature selection such Relief [4] or FSDD [6]. Feature extraction reduces the dimensionality by (linear or non-linear) projection of Q-dimensional vector on to q-dimensional vector (q<<Q). However, it changes the original physical features and makes features uninterpretable. Feature selection reduces the dimensionality by selecting a subset of original variables. Features selection methods tend to produce less expensive classifications, the non-selected variables are not longer needed and they are more easily interpretable.

The goal of feature selection is to find the optimal subset of q features chosen from the total Q features. An exhaustive search strategy to seek the best subset among all the possible Q over q feature subsets usually results in a considerably high computational effort. In order to simplify this complex problem, most of the existing feature selection methods convert the problem into a feature ranking problem. The algorithm finds out the features that promise good class separability among different classes as well as keep the samples in the same class as close as possible.

For this article we have selected two well-known algorithms susceptible to handle interval attributes: Relief and FSDD. Relief is a widespread feature selection algorithm, probably due to its simplicity and good results. FSDD has as main advantage a proof of optimality of its ranking strategy.

2.1 FSDD

Feature selection based on distance discriminant (FSDD), proposed by Liang et al. [6], select features by maximizing the criterion:

$$d_b - \beta \cdot d_w \tag{1}$$

where d_b is the distance between different classes, d_w corresponds to the distance within classes, and the parameter β is used to control the balance between d_b and d_w. As it is proved in [6] by using suitable definitions of d_b and d_w, the value of (1) can be transformed into the form:

$$d_b - \beta \cdot d_w = \sum_{k=1}^{m} \frac{1}{\sigma_k^2} \left[\sigma''^2_k - \beta \sum_{i=1}^{c} \rho_i \cdot \sigma'^2_k(i) \right] \tag{2}$$

where m is the number of selected features, c the number of classes, ρ_i the prior probability of i-th class, σ_k^2 the standard deviation in k-th feature, $\sigma'^2_k(i)$ the standard deviation in i-th class (having n_i samples) in k-th feature, and σ''^2_k the weighted standard deviation of the class center in the k-th feature (see [6] for more details).

According to equation (2) the optimal feature subset can be chosen by ranking the features in descending order according to the function:

$$f(k) = \frac{1}{\sigma_k^2} \left[\sigma''^2_k - \beta \sum_{i=1}^{c} \rho_i \cdot \sigma'^2_k(i) \right] \tag{3}$$

Then the optimal subset of m features maximizing $f(k)$ is the first m features sorted by the feature ranking.

2.2 ReliefF

Kira and Rendell [4] developed the algorithm called Relief based on attribute estimation. However, the original Relief cannot deal with incomplete data and it was limited to two-class problems only. Kononenko et al. [5] developed an extension of Relief, called ReliefF, which improved the original algorithm and extended it to handle incomplete and multi-class data sets while the complexity remains the same.

The ReliefF algorithm (see figure 2) randomly selects an instance R_i and then searches for k of its nearest neighbors from the same class, called nearest hits H_j and also k nearest neighbors from each of the different classes, called nearest misses $M_j(C)$. It updates the quality estimation $W(a)$ for all attributes a depending on their values for R_j, hits H_j and misses $M_j(C)$. In the update formula, ReliefF averages the contribution of all the hits and all the misses. Moreover, the contribution for each class of the misses is weighted with the prior probability of that class $P(C)$.

The $diff(a,R_i,R_j)$ function computes how different the values for feature a are in examples R_i and R_j,

$$diff(a, R_i, R_j) = \frac{|value(a, R_i) - value(a, R_j)|}{max(a) - min(a)} \tag{4}$$

where $value(a,R_i)$ denote the value of feature a on example R_i, and $max(a)$ and $min(a)$ the maximum and minimum value for a, respectively.

```
Alforithm ReliefF
Input: for each training instance a vector of attribute
values and the class value.
Output: the vector W of estimations of the qualities of
attributes.
set all weights W[a]:=0.0;
for i:=1 to m do begin
     randomly select an instance Ri
     find k nearest hits Hj;
     for each class C≠class(Ri) do
```

```
find k nearest misses Mj(C);
for a:=1 to Q do
```

$$W[a] := W[a] - \sum_{j=1}^{k} \frac{diff(a,R_i,H_j)}{m \cdot k} + \sum_{C \neq class(R_i)} \left[\frac{P(C)}{1 - P(class(R_i))} \sum_{j=1}^{k} \frac{diff(a,R_i,M_j(C))}{m \cdot k} \right]$$

```
end
```

Figure 2. Pseudo code of ReliefF algorithm

3 Kernels in interval-valued data

A kernel defined in a set A is a bi-valued one-dimensional function k from $A \times A$ to R that for all set $\{a_1,\ldots a_n\}$ with arbitrary $n \in N$, matrix $K=[k(a_i,a_j)]$ (Gramm matrix) is symmetric and positive semidefinite. It has been demonstrated that for all function k which satisfies this property, there exists a Hilbert space $\{F, \langle\rangle\}$ and a map Φ from A to F verifying $k(a_i,a_j)= \langle\Phi(a_i), \Phi(a_j)\rangle$, i.e., the result of applying the function k to the pair of elements (a_i, a_j) from $A \times A$ is equivalent to calculate the dot product of the images by the map Φ. On the other side, if a map Φ from A to F can be found, where F is a Hilbert space, the function $k(a_i,a_j)= \langle\Phi(a_i), \Phi(a_j)\rangle$ is a kernel [3].

Kernels were first introduced in the Support Vector Machines (SVM) approach with the aim of using this machine learning method with non-linearly separable data. Due to the advantage of not requiring any Euclidean structure in the input space but a Hilbert space structure, SVM (and other kernalizable learning methods) can be used with patterns described by variables not belonging to any Euclidean space, for instance patterns described by intervals.

3.1 Intersection kernel

The set of the intervals do not directly have a Euclidean structure. To define kernels in the set of the intervals, however, is possible. A first kernel to be defined in the set of the intervals is the length of the intersection kernel, $K_\cap$.

Theorem

Let $I_1=[a,b]$, $I_2=[c,d]$ are two real intervals, then a map defined as:

$$K_\cap(I_i, I_2) = length(I_1 \cap I_2) \tag{5}$$

is a kernel.

Proof

Let Φ be a map that associates to each interval $I=[a,b]$ an indicator function ϕ_I from the Hilbert space L_2 of square integrable real functions defined on $(-\infty, +\infty)$ of the following way:

$$\phi_{[a,b]}(x) = \begin{cases} 1 & if \quad a \le x \le b \\ 0 & otherwise \end{cases} \tag{6}$$

The composition of this function with the usual dot product in L_2 leads to the function $K_\cap$.

$$K_\cap([a,b],[c,d]) = \langle \phi_{[a,b]}, \phi_{[c,d]} \rangle$$
$$= \int_{-\infty}^{+\infty} \phi_{[a,b]}(x) \cdot \phi_{[c,d]}(x) \cdot dx = length([a,b] \cap [c,d])$$

Therefore, $K_\cap$ is a kernel (see [3])

■

3.2 Influence functions

One of the disadvantages of the intersection kernel defined above is that it is unable to discriminate between disjoint intervals more or less separated, since result is ever null. This fact is particularly serious if the length of the intervals is very small.

A modification of the intersection kernel that takes into account not only the common part but also the relative distance between them is the intersection kernel with exponential influence introduced in [7]. This kernel is based on an influence function that extends the range of influence beyond the range of the interval.

$$f_{[a,b],\sigma}(x) = \begin{cases} e^{-\frac{a-x}{\sigma}} & si \quad x < a \\ 1 & si \quad a \le x \le b \\ e^{-\frac{x-b}{\sigma}} & si \quad x > b \end{cases} \tag{7}$$

From this exponential influence function the kernel is defined as:

$$K([a,b],[c,d]) = \int_{-\infty}^{+\infty} f_{[a,b]}(x) \cdot f_{[c,d]}(x) \cdot dx \tag{8}$$

3.3 A kernel as a similarity measure

The simplest kernel defined on a Euclidean space, i.e. the dot product, is a non-normalized measure of similarity. The quotient between the dot product and the product of the vector norms, known as the cosine of the angle, is a number between -1 and 1, by virtue of the Cauchy-Schwarz inequality. This result can be generalized to more general positive semidefinite kernels

Theorem (Cauchy- Schwarz inequality for kernels). *If $K{:}A\times A\rightarrow R$ is a kernel (symmetric and positive semidefinite) it is verified the inequality:*

$$-1 \le \frac{K(a_i,a_j)}{\sqrt{K(a_i,a_i)\cdot K(a_j,a_j)}} \le 1 \tag{9}$$

If, in addition, the kernel is a positive function, i.e. $K{:}A\times A\rightarrow R+$, it is verified:

$$0 \le \frac{K(a_i,a_j)}{\sqrt{K(a_i,a_i)\cdot K(a_j,a_j)}} \le 1 \tag{10}$$

Proof

Since K is positive semidefinite, 2×2 Gramm matrix $K_{ij}=K(a_i,a_j)$, $i,j\in\{1,2\}$ is also positive semidefinite, therefore the determinant $K_{11}\cdot K_{22}-K_{12}^{\;2}\ge 0$, that is:

$$\frac{K_{12}^{\;2}}{K_{11}\cdot K_{12}} \le 1 \Rightarrow -1 \le \frac{K_{12}}{\sqrt{K_{11}\cdot K_{12}}} \le 1 \qquad \blacksquare$$

3.4 Generalization to multidimensional intervals

A q-D generalized interval is the Cartesian product of q reals intervals. Its representation is a hyperrectangle or q-orthotope. The most direct way of defining a kernel on the set of multidimensional intervals is by using the product of kernels of its components. If the intersection kernel is used, this product can be interpreted as the hypervolume of the intersection. However, another way to define a kernel in the set of multidimensional intervals is to define the kernel as:

$$K(I_i,I_j) = \sum_{k=1}^{q} \frac{K(I_{ik},I_{jk})}{\sqrt{K(I_{ik},I_{ik})\cdot K(I_{jk},I_{jk})}} \tag{11}$$

This function is a kernel due to the sum of kernels is also a kernel [3]. Moreover, the sum is well defined, because the summands are dimensionless. The expression of the multi-dimensional interval kernel as a sum of normalized kernel of its components will allow to easily generalize the FSDD method to interval features.

4 Feature Selection with interval-valued data

Let us start by introducing the notation used in this section. Let us assume that there exist a set of $N=n_1+n_2+\ldots+n_C$ patterns belonging to C classes (n_i patterns belonging to i-th class) and each pattern characterized by Q interval features: $I_j^{\,i}=(I_{j1}^{\,i},\ldots, I_{jQ}^{\,i})$, where $I_{jk}^{\,i}=[c_{jk}^{\,i}-r_{jk}^{\,i}, c_{jk}^{\,i}+r_{jk}^{\,i}]$ represents the k-th feature of j-th pattern belonging to class i-th which has center $c_{jk}^{\,i}$ and radius $r_{jk}^{\,i}$.

4.1 Interval FSDD

In order to extend the FSDD algorithm to handle with interval variables, first, distances are substituted by similarity measures, i.e. equation (1) are substituted by:

$$S_b - \beta\cdot S_w \tag{12}$$

where S_b is a similarity measure between classes and S_w is a similarity measure within classes. The parameter β is used to control the balance between S_b and S_w .

Let us consider the multidimensional mean interval $m^i=(m_1^{\,i}, m_2^{\,i},\ldots,m_Q^{\,i})$ representing the whole i-th class being $m_k^{\,i}=[c_k^{\,i}-r_k^{\,i}, c_k^{\,i}+r_k^{\,i}]$ and:

$$\begin{aligned} c_k^i &= \frac{1}{n_i}\sum_{l=1}^{n_i} c_{kl}^i \\ r_k^i &= \frac{1}{n_i}\sum_{l=1}^{n_i} r_{kl}^i + \sqrt{\frac{1}{n_i}\sum_{l=1}^{n_i}\left(c_{kl}^i - c_k^i\right)^2} \end{aligned} \tag{13}$$

i.e., the center of the mean interval is the mean of centers and the radius is the mean of radius plus the standard deviation of centers.

Definition 1. (Similarity between classes)

If $S(m^i,m^j)$ is some similarity measure between multidimensional mean intervals m^i and m^j from classes i and j, and $\rho_i=n_i/N$ the prior probability of the i-th class, we define de similarity between classes as:

$$S_b = \sum_{i=1}^{c}\rho_i \sum_{j>i}^{c}\rho_j \cdot S(m^i,m^j) \tag{14}$$

In our work we consider the interval kernel defined above as a similarity measure between intervals, that is:

$$\begin{aligned} S_b &= \sum_{i=1}^{c}\rho_i \sum_{j>i}^{c}\rho_j \cdot K(m^i,m^j) = \\ &= \sum_{i=1}^{c}\rho_i \sum_{j>i}^{c}\rho_j \sum_{k=1}^{Q} \frac{K(m^i{}_k,m^j{}_k)}{\sqrt{K(m_k^i,m_k^i)\cdot K(m_k^j,m_k^j)}} = \\ &= \sum_{k=1}^{Q}\left(\sum_{i=1}^{c}\rho_i \sum_{j>i}^{c}\rho_j \frac{K(m^i{}_k,m^j{}_k)}{\sqrt{K(m_k^i,m_k^i)\cdot K(m_k^j,m_k^j)}} \right) \end{aligned} \tag{15}$$

Definition 2. (Similarity within classes)
We define the term associated to the similarity within classes as:

$$S_w = \sum_{i=1}^{c} \rho_i \frac{2}{n_i \cdot (n_i - 1)} \sum_{j=1}^{n_i} \sum_{l>j}^{n_i} K(I_j^i, I_l^i) \tag{16}$$

It is derived from definition 1 and 2 and equation (11) that:

$$\begin{aligned} S_w &= \sum_{i=1}^{c} \rho_i \frac{2}{n_i \cdot (n_i - 1)} \sum_{j=1}^{n_i} \sum_{l>j}^{n_i} K(I_j^i, I_l^i) = \\ &= \sum_{i=1}^{c} \rho_i \frac{2}{n_i \cdot (n_i - 1)} \sum_{j=1}^{n_i} \sum_{l>j}^{n_i} \sum_{k=1}^{Q} \frac{K(I_{jk}^i, I_{lk}^i)}{\sqrt{K(I_{jk}^i, I_{jk}^i) \cdot K(I_{lk}^i, I_{lk}^i)}} = \\ &= \sum_{k=1}^{Q} \left(\sum_{i=1}^{c} \rho_i \frac{2}{n_i \cdot (n_i - 1)} \sum_{j=1}^{n_i} \sum_{l>j}^{n_i} \frac{K(I_{jk}^i, I_{lk}^i)}{\sqrt{K(I_{jk}^i, I_{jk}^i) \cdot K(I_{lk}^i, I_{lk}^i)}} \right) \end{aligned} \tag{17}$$

and:

$$\begin{aligned} S_b - \beta \cdot S_w = \sum_{k=1}^{Q} \Bigg(& \sum_{i=1}^{c} \rho_i \sum_{j>i}^{c} \rho_j \frac{K(m^i{}_k, m^j{}_k)}{\sqrt{K(m_k^i, m_k^i) \cdot K(m_k^j, m_k^j)}} - \\ & - \beta \cdot \sum_{i=1}^{c} \rho_i \frac{2}{n_i \cdot (n_i - 1)} \sum_{j=1}^{n_i} \sum_{l>j}^{n_i} \frac{K(I_{jk}^i, I_{lk}^i)}{\sqrt{K(I_{jk}^i, I_{jk}^i) \cdot K(I_{lk}^i, I_{lk}^i)}} \Bigg) \end{aligned} \tag{18}$$

According to the equation (18) the optimal feature subset can be chosen by ranking the Q features in ascending order according to the evaluation function:

$$\begin{aligned} & \sum_{i=1}^{c} \rho_i \sum_{j>i}^{c} \rho_j \frac{K(m^i{}_k, m^j{}_k)}{\sqrt{K(m_k^i, m_k^i) \cdot K(m_k^j, m_k^j)}} - \\ & - \beta \cdot \sum_{i=1}^{c} \rho_i \frac{2}{n_i \cdot (n_i - 1)} \sum_{j=1}^{n_i} \sum_{l>j}^{n_i} \frac{K(I_{jk}^i, I_{lk}^i)}{\sqrt{K(I_{jk}^i, I_{jk}^i) \cdot K(I_{lk}^i, I_{lk}^i)}} \Bigg) \end{aligned} \tag{19}$$

Then the optimal subset of q features minimizing equation (19) is the first q features sorted by the feature ranking. This is due to the fact that expressions of S_w and S_b are a sum of expressions depending on each feature.

4.2 Interval ReliefF

In order to handle with interval features in ReliefF, also the similarity measure introduced by the interval kernel is employed. Firstly, the hits and misses are selected by maximizing the kernel between the pattern considered and the rest of patterns from the same class (hits) or other classes (misses).

On the other hand, the function *diff*() that appears in the original algorithm is changed by the interval version:

$$diff(k, R_i, R_j) = \frac{K(R_{ik}, R_{jk})}{\sqrt{K(R_{ik}, R_{ik}) \cdot K(R_{jk}, R_{jk})}} \tag{20}$$

where R_{ik} and R_{jk} are the *k-th* features of patterns R_i and R_j respectively.

5 Example

In order to evaluate the introduced feature selection methods and interval data abilities in signal recognition tasks, a comparison using intervals and data points is performed. A learning algorithm will be tested using features selected by ReliefF and FSDD methods when using standard single data. Same learning algorithm adapted to interval kernels will be evaluated using those attributes selected by the interval version of the feature selection methods.

The 'Character Trajectories' dataset [8] belonging to the UCI Machine Learning Repository [1] will be employed for the proposed evaluation. Samples of this dataset represent three signals captured while handwriting characters on a digital tablet. Signals are (x, y) position over the tablet and pen tip force along time. The number of samples in this dataset is 2858 divided in 20 classes, as many as possible characters.

A pattern is described as a set of interval values that represents its signals. An interval is extracted from the first and third quartile of a data variable. This representation is used because it is less sensitive to outliers than using the maximum and minimum. Variables used to obtain intervals are the original signals and new ones, which are created from the original through discrete derivatives, proportion between the m and its norm. Taking into account all variables, it is considered 26 interval-valued features, meaning that each handwritten character is represented by a pattern comprised of 26 intervals. In order to obtain results that can be fairly compared, standard single point features are the mean and the deviation of each variable; hence 52 features are considered.

5.1 Experiments setup

Support Vector Machines are chosen as the learning algorithm because it is able to handle both, interval-valued and standard single point-data, if suitable kernels are used. An interval kernel with influence functions is used in feature selection and classification tasks when using interval data. Standard FSDD and ReliefF algorithms are used for regular data, as well as a RBF kernel when classifying.

Libsvm library [2] and multiclass one-vs-one approach are used. Interval kernel parameter σ value is fixed to 1 in feature selection tasks. Its effect on classification performance is tested using values 2^3, 2^2… 2^{-2}, 2^{-3}, and 0 (without influence function). Regularization parameter C existing on both classification methods and RBF kernel parameter γ are selected by stratified 5-fold Cross-Validation (CV) using values from 2^{-3}, 2^{-2}… 2^2, 2^3 and 10^{-3}, 10^{-2}… 10^2, 10^3 respectively.

Accuracy is defined as the proportion of correctly classified samples over total number of samples. Each classification process is performed 10 times in order to take into account CV variability, so mean accuracy along repetitions is considered in results. A tenth part of each class of the dataset is used in order to obtain results in a feasible time.

5.2 Results and discussion

Classification accuracies using best n features given by each feature selection algorithm are obtained, with n=1..10. Mean accuracies and its deviation between repetitions are showed in table 1. Results obtained by standard data using ReliefF are similar to those obtained by its interval-valued version. FSDD on RBF provides better results than FSDD-I, although both achieve similar performances with 10 features. ReliefF provides lower accuracies than FSDD in both kernel types.

Nº feats.	Intersection kernel		RBF kernel	
	FSDD-I	**ReliefF-I**	**FSDD**	**ReliefF**
1	24.1 ±7.6	24.8 ±6.7	25.6 ±3.3	30.8 ±3.1
2	40.6 ±8.9	54.3 ±8.5	60 ±3.6	60.9±3.8
3	50.3 ±10.8	62.8 ±6	72.4 ±3.2	66.±4.2
4	52.1 ±8.3	66.3 ±7.8	78.4 ±2.7	71.6±3.7
5	55.7 ±6.6	68.6 ±10.5	82 ±2.7	73.1 ±4.2
6	66.1 ±7.9	73 ±7.1	83.2 ±3.1	74.5±3.7
7	69.9 ±8.2	74.3 ±7.3	82.8 ±3.2	75.9±4.7
8	74.7 ±6	77 ±7.5	83.9 ±3	76.7 ±3.9
9	77.6 ±8.1	77.9 ±6.7	84.2 ±3.5	79.2±3
10	79 ±6.9	77.9 ±5.5	83.5 ±3.3	81.9 ±4.1

Table 1. Mean accuracy and deviation between repetition results

Intersection kernel results do not outperform the ones using RBF kernel. Instead, similar values are obtained. This means that interval valued data may be a reliable way to represent and classify signals, as much as standard features in the dataset used. The advantage would be the simpler and more understandable form the data is represented.

6 Conclusion and Future work

This work proposes a combination of feature selection algorithm and kernel-based classification in order to perform signal recognition based on interval-valued data. To this end, an intersection-kernel that employs influence functions is presented and used as a similarity measure in interval-valued versions of two standard feature selection algorithms. The kernel is also used to perform classification, and its results using features suggested by interval versions of feature selection algorithms are compared to the obtained by using a RBF kernel on features indicated by the original algorithms with standard single point data. Results show that similar accuracies between them are obtained, suggesting that interval values are a reliable representation of data in signal recognition tasks.

As future work, we are planning to apply the methodology on new datasets, including artificial data, in order to detect which particular cases of signal recognition interval-valued algorithms can provide results with higher or lower accuracy. Moreover, we want to test its abilities on aggregating large amount of data and its possibilities on handling uncertain data.

References

[1] Asuncion, A. & Newman, D.J. *UCI Machine Learning Repository:* http://www.ics.uci.edu/~mlearn/MLRepository.html. Irvine, CA: University of California, School of Information and Computer Science. 2007

[2] Chih-Chung Chang and Chih-Jen Lin, LIBSVM: a library for support vector machines, 2001. Software available at *http://www.csie.ntu.edu.tw/~cjlin/libsvm*

[3] Cristianini, N and Shawe-Taylor, J. *Support Vector Machines and other kernel-based learning methods*. Cambridge University Press,

[4] Kira, K. & Rendell, L.A. A practical approach to feature selection. *Machine Learning* (pp. 249-256), 1992

[5] Krononenko, I. Estimating attributes: analysis and extensions of relief. *ECMS-94: Proceedings of the European conference on machine learning on Machine Learning* (pp 171-182). 1994

[6] Liang J, Yang S. & Winstanley A. Invariant optimal feature selection: A distance discriminant and feature ranking based solution. *Pattern Recognition* vol 41, Issue 5, (pp 1429-1439). 2008

[7] Ruiz, F.J. *Funciones núcleo sobre estructuras cualitativas*. PhD dissertation. Universitat Politècnica de Catalunya. 2009

[8] Williams, B.H, Toussaint, M. and Storkey , A.J. Modelling motion primitives and their timing in biologically executed movements. In *J.C. Platt, D. Koller, Y. Singer, and S. Roweis, editors, Advances in Neural Information Processing Systems* 20 , (pp 1609–1616). MIT Press , Cambridge, MA, 2008.